Best Practices –Preparation of Data for Inclusion in GIS

A number of data types can be loaded into a GIS project, but only a few are recommended. Data standards must be met for the information within these files before data will function properly within GIS software.

File Types

Preferred file types are tab delimited text (TXT), dBase (DBF), or Access tables. Excel files are spreadsheets, not databases; they must be converted to one of these three types before they can be used in GIS. Excel files will usually need to be edited to conform to the required standards. After the edits have been made, then the Excel file can be saved as either TXT or DBF. Text (TXT) formatted files are less problematic.

Data that was copied from a web page

Data that was copied from a web page into Excel should always be saved in the TXT file format. This will remove any formatting (HTML) that is still associated with the data.

File Preparation - Excel

Here are some guidelines for making your Excel file 'GIS ready'. This will prepare your Excel file for conversion to either TXT or DBF format.

- Keep a copy of your original file.
- The first line must contain ONLY the names of the fields.
- Use only one line for your column headings, don't create 'stacked' headings
- There should be NO additional header, footer or sidebar comments in the document. This includes summary lines, totals, references, citations, etc.
- Don't use punctuation characters (- & % # / etc) for either field names or within the data. There is one exception; you can use _ (underscore).
- Headings should start with a letter, not a number.
  Example: 2002rate should be changed to rate02 or rate2002
- If you intend to make DBF files, rename your headings to be as brief as possible. For example 2002 rates / 10000 would become Rate02. Maximum field length for DBF is 10. (It will be truncated if it's too long.) The following field names will no longer be unique when truncated at the 10th character:

  DHSSGIS_STREETNAME
  DHSSGIS_STREETTYPE
  DHSSGIS_STATE

  NOTE: The 10 character limit is not a restriction in TXT formats.
- Remove spaces between words. Instead, use capital letters to distinguish words.
  Example: PctTotal02
- Field names should be in all UPPER case or all mixed/proper; either is OK, as long as it is consistent.

- Remain consistent with your capitalization. If you have multiples files, they should contain data in all UPPER case (e.g. COLE, OSAGE, BOONE) or all mixed/proper case (e.g. Cole, Osage, Boone).
- Do not use 1000 separators in your columns of numbers. This can be changed under Format Cells > Number Uncheck 'Use 1000 separator (,)'.
- For whole numbers, set the number of decimal places to 0. This can be changed under Format Cells > Number. Use the down arrow next to number found in 'Decimal places:' or key in '0'.
- For character/text data, make sure there are no leading blanks. Text should appear justified to the left, without spaces before the first character.
- **Bold** text, boxes around cells and other cosmetics can be left as-is. This formatting will be stripped off automatically during the conversion from Excel to the other formats.

For files of county data:

- Suggested field name for county names is County or COUNTY.
- Make sure there are no misspellings in your county names. The most common misspellings are:
  Daviess – sometimes has the 2nd 'S' missing
  DeKalb, McDonald – The 'K' and 'D' should be capitalized
  Ste. Genevieve – sometimes is missing the 'E' in Ste.
- Do not include the word 'County' in the name of the county.
- For St. Louis – it is St. Louis and St. Louis City – you should not add the word 'County' after St. Louis to denote the county of St. Louis.
- All of the county names that start with Saint should be abbreviated as St. (be sure and get the period at the end). The abbreviation for Sainte Genevieve is Ste. (feminine form of the saint abbreviation).
- If you have the Federal Information Processing Standards or FIPS code, this may be used instead of the county name.
- St. Louis City is equivalent to a county for Federal and State government purposes. St. Louis City is both a city and a county. Kansas City is NOT a county or equivalent. If possible, create a separate field to contain the reporting Jurisdication rather than using the county column. Kansas City actually falls within four separate counties. (Independence and Joplin are often separate reporting jurisdictions from their counties and should be treated like Kansas City, as a jurisdiction and not a county if possible.)

For files of ZIP code data:
ZIP codes should be stored as text field types. The same goes for any ID code that has only numbers and may have a zero as the first character, such as three digit county FIPS codes. The leading zero will be lost if the field is stored as a numeric field (integer, float, single or double).

Example

BEFORE

	A	B	C	D	E	F	G	H	I
1	AGE OF THE POPULATION 60 YEARS AND OVER [100]						*Source: U.S. Census - 2000*		
2		Universe: Population 60 years and over							
3	[Based on a sample. Rounded data. Geographic areas are omitted if they do not meet the specified population threshold.]								
4	Geographic Name	Total:	60 to 64 years:		65 to 84 years:		85 years and over:		
5			Total	Percent of ≥ 60 Population	Total	Percent of 65 - 84 Population	Total	Percent of ≥ 85 Population	
6									
9	Adair	3850	800	0.2078	2590	0.6727	460	0.1195	
10	Andrew	3000	630	0.2100	2070	0.6900	300	0.1000	
11	Atchison	1640	290	0.1768	1090	0.6646	260	0.1585	
12	Audrain	5570	1130	0.2029	3870	0.6948	560	0.1005	
13	Barry	7280	1800	0.2473	4860	0.6676	610	0.0838	
14	Barton	2650	580	0.2189	1770	0.6679	300	0.1132	

AFTER

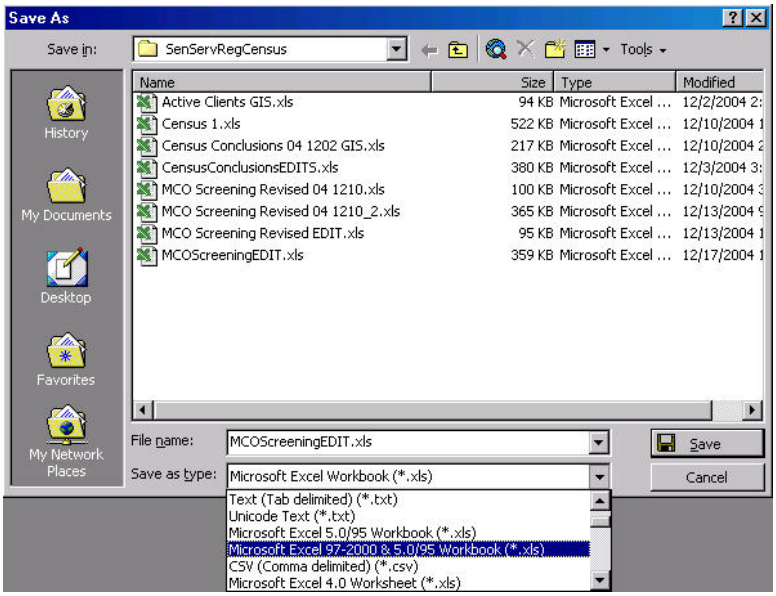
	A	B	C	D	E	F	G	H
1	COUNTY	TOTPOP	AGE60_64	PCT60_64	AGE65_84	PCT65_84	OVER85	PCTOVER85
2	ADAIR	3850	800	0.2078	2590	0.6727	460	0.1195
3	ANDREW	3000	630	0.2100	2070	0.6900	300	0.1000
4	ATCHISON	1640	290	0.1768	1090	0.6646	260	0.1585
5	AUDRAIN	5570	1130	0.2029	3870	0.6948	560	0.1005
6	BARRY	7280	1800	0.2473	4860	0.6676	610	0.0838
7	BARTON	2650	580	0.2189	1770	0.6679	300	0.1132
8	BATES	3690	790	0.2141	2510	0.6802	390	0.1057

IMPORTANT: Save your Excel file first, before saving to another format.

After saving your file in its Excel format, then you can save it as a text (TXT) or dBase file (DBF 4). Under the File menu, select Save As.

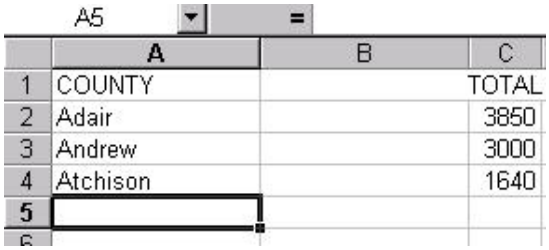
If your Excel file contains multiple worksheets, you will have to save each worksheet into a separate file. Be aware than once you have saved to the new file type, you will have to re-open the Excel file again to save any other worksheets.

Files cannot be saved to Access tables from Excel, but Access can import existing Excel files.



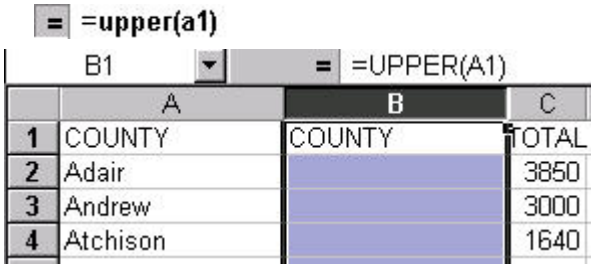
How to convert text to all upper case or mixed case - Excel

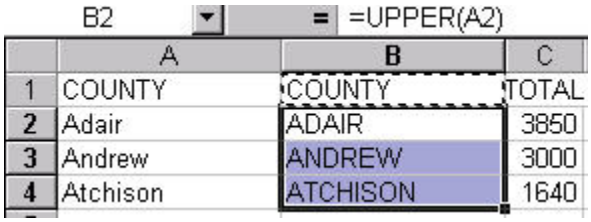
Converting data, such as county names, to all upper case is one way to create consistency in your data set. This conversion can be accomplished with a function in Excel.



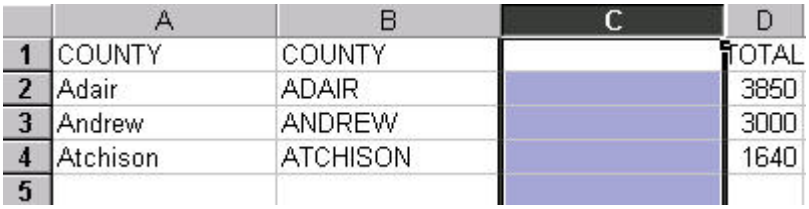
Create a new empty column next to the column you need to convert. Select the column to the left of the column you need to convert, click Insert > Columns. A column will be added to the file.

Create a function for this new column. Select the first empty cell at the top of the new column. Use the = sign (next to the place where you can type at the top). Type the function:





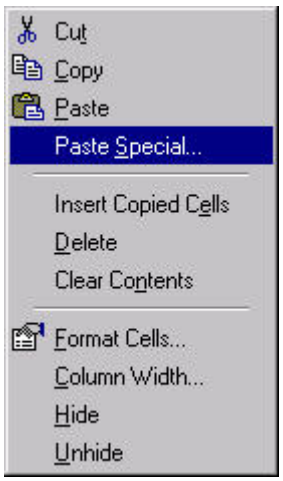
Copy and paste this function from the first cell into the rest of the cells in the empty column. This will change the format of the data to all upper case. This column will only exist temporarily.



Create another empty column next to the column you just created.

You will now copy the column containing the calculated functions to this new empty column. Right-click on the column heading of the calculated function (in the example, column B), select Copy.

Right click on the new empty column (in the example, column C) and Select Paste Special.



You only want to copy the values for the county names, not the formula used to create them.

Make sure to click the choice for Values in the dialog. (This is not the default.)

C2 = ADAIR

	A	B	C	D
1	COUNTY	COUNTY	COUNTY	TOTAL
2	Adair	ADAIR	ADAIR	3850
3	Andrew	ANDREW	ANDREW	3000
4	Atchison	ATCHISON	ATCHISON	1640

This will look identical to you, and visually they are the same. The difference is that one column (column B in this case) contains a function and not the actual data. Column C contains only the text values.

	A	B	C	D
1	COUNTY	COUNTY	COUNTY	TOTAL
2	Adair	ADAIR	ADAIR	3850
3	Andrew	ANDREW	ANDREW	3000
4	Atchison	ATCHISON	ATCHISON	1640
5				

To complete the conversion, you must delete the original unformatted column and the temporary column.
Select the original column and the temporary column. Select Edit > Delete.

	A	B
1	COUNTY	TOTAL
2	ADAIR	3850
3	ANDREW	3000
4	ATCHISON	1640
5		

This will leave only one column of properly formatted values.

If you need to convert UPPER case to Proper case, the steps are very similar. Use the function Proper rather than Upper, and otherwise follow the same steps above.

When converting from UPPER to Proper, you must go back and fix DeKalb, McDonald, making sure the 'K' and the 'D' are capitalized.

File Preparation – Access

Here are some guidelines for making your Access table 'GIS ready':

- Tables in Access should have a field (suggested name ID) defined as an AutoNumber and set as the primary key.
- Don't use punctuation characters (- & % # / etc). There is one exception; you can use _ (underscore).
- Headings should start with a letter, not a number.
  Example: 2002rate should be changed to rate02 or rate2002
- Field names should be in all UPPER case or all mixed/proper; either is OK, as long as it is consistent.
- Remain consistent with your capitalization. If you have multiples tables, they should contain data in all UPPER case (e.g. COLE, OSAGE, BOONE) or all mixed/proper case (e.g. Cole, Osage, Boone).
- If the data contains coordinates, make sure these fields are defined as Number, with a field size of Double.
- There are several field names that are reserved and should not be used in Access. They can be found at: http://support.microsoft.com/default.aspx?scid=kb;EN-US;q248738

For files of county data:

- Suggested field name for county names is County or COUNTY.
- Make sure there are no misspellings in your county names. The most common misspellings are:
  Daviess – sometimes has the 2nd 'S' missing
  DeKalb, McDonald – The 'K' and 'D' should be capitalized (in mixed case)
  Ste. Genevieve – sometimes is missing the 'E' in Ste.
- Do not include the word 'County' in the name of the county.
- For St. Louis – it is St. Louis and St. Louis City – you should not add the word 'County' after St. Louis to denote the county of St. Louis.
- All of the county names that start with Saint should be abbreviated as St. (be sure and get the period at the end). The abbreviation for Sainte Genevieve is Ste. (feminine form of the saint abbreviation).
- If you have the Federal Information Processing Standards or FIPS code, this may be used instead of the county name.

- St. Louis City is equivalent to a county for Federal and State government purposes. St. Louis City is both a city and a county. Kansas City is NOT a county or equivalent. If possible, create a separate field to contain the reporting Jurisdication rather than using the county column. Kansas City actually falls within four separate counties. (Independence and Joplin are often separate reporting jurisdictions from their counties and should be treated like Kansas City, as a jurisdiction and not a county if possible.)

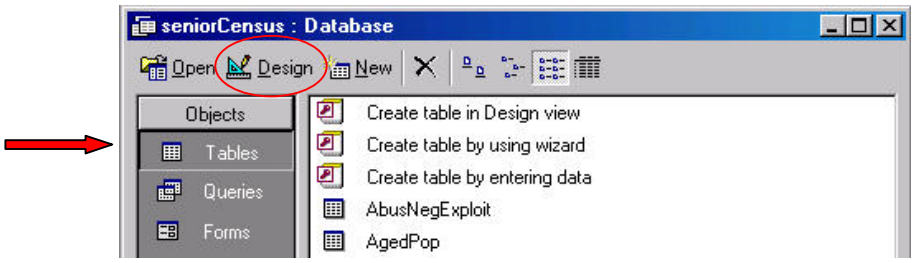
For files of ZIP code data:
ZIP codes should be stored as text field types. The same goes for any ID code that has only numbers and may have a zero as the first character, such as three digit county FIPS codes. The leading zero will be lost if the field is stored as a numeric field (integer, float, single or double).

Access tables do not need to be converted to any other file type. Send the entire MDB file for inclusion with GIS projects.

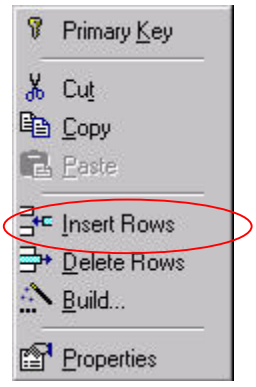
How to convert text to all upper case or mixed case – Access

If you need to convert your county names, or other data to all upper case to create consistency, this can be accomplished with an update query in Access.

Add a new field for your modified data.

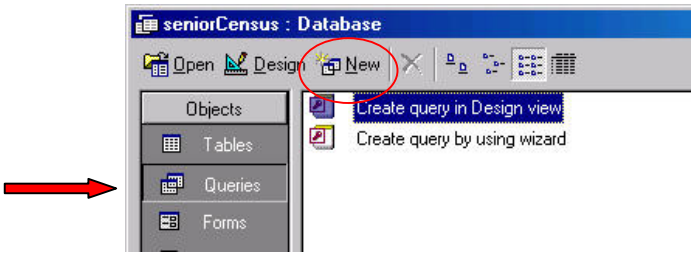


Open the Table in Design View

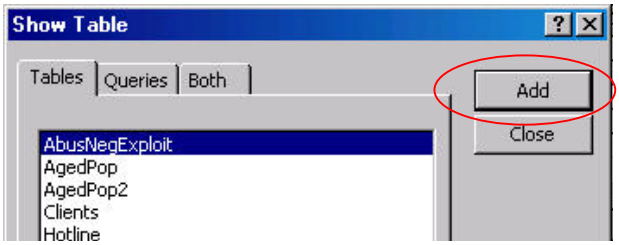


Add a new row under the field name you wish to modify. Suggested field name is <originalName>2 (just add a '2' to the end of the field name). Save your changes and close the Design View.

Select the Queries option and then select New.



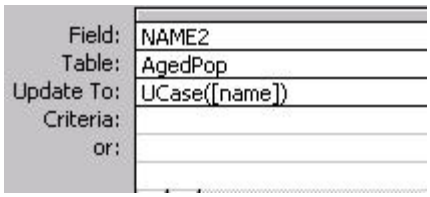
Select the table name from the list of available tables and click Add.



Be default, all new queries in Access are Select Queries. You must change from a Select Query to an Update Query. Click on the down arrow next to the Select Query icon and select Update Query (shown as a pencil).

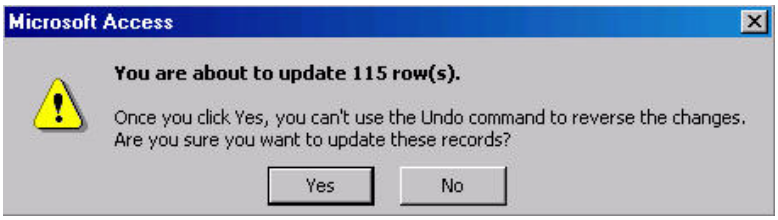
Click the new field you created and drag it to the dialog at the bottom of the Query window. This will add the field name with the source table name below it.

In the next row, type in the function UCase([myfieldname2]), substituting your field name for myfieldname2.



Click the Run icon. 

A dialog about updating ## rows will appear. Select Yes.



There is no Access function to convert upper case text to mixed/proper case. You can export the table to Excel in order to use the function Proper, described in the previous section ‘How to convert text to all upper case or mixed case – Excel’.